DOCUMENT RESUME

ED 214 168 CS 206 776

AUTHOR Quellmalz, Edys; And Others

TITLE Defining Writing: Effects of Discourse and Response Mode.

INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.

SPONS AGENCY National Inst. of Education (ED), Washington, D.C.
REPORT NO CSE-R-132
PUB DATE 80
NOTE 59p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Discourse Analysis; Educational Assessment; Holistic Evaluation; *Measurement Techniques; *Responses; Secondary Education; *Test Format; *Writing Evaluation; *Writing Research

ABSTRACT

The complexity of writing as a skill domain and the lack of consensus about its components have engendered much controversy about the type, length, or number of tasks that should be administered in a given test form and even about whether some aspects of composition require direct assessment through writing samples. Acknowledging this, a study was conducted to examine the comparability of writing competency profiles derived from test tasks differing in discourse and response mode. Two hundred high school students were given a multiple choice test and a paragraph writing task, as well as two full-length essay assignments. Ratings of the essays and paragraph on an analytic scale and scores on the objective test provided the bases for comparisons. Results indicated that levels of performance varied on tasks presenting different writing purposes. Also shown was that repeated applications of the scoring rubric produced measures that taped the same underlying content. In addition, factors reflecting the content of the writing subscale were strongly intercorrelated, a factor which is present no matter what response mode the subjects were assessed in. (HOD)

DEFINING WRITING:
EFFECTS OF DISCOURSE AND RESPONSE MODE

Edys Quellmalz, Frank J. Capell,
and Chih-Ping Chou

CSE Report No. 132
1980

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California, Los Angeles

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

EFFECTS OF DISCOURSE AND RESPONSE MODE ON THE MEASUREMENT OF WRITING COMPETENCE

As school district and state assessment programs attempt to test student basic skills achievement, attention to the methodological problems inherent in measuring writing competence increases. The complexity of writing as a skill domain and the lack of consensus about its components have engendered much controversy about the type, length or number of tasks that should be administered for a given test form and even about whether some aspects of composition require "direct" assessment through the elicitation of writing samples.

Two salient measurement issues involved in specifying writing task types are the response mode (selected vs. constructed) and the discourse mode required by the tasks. Conventionally, large scale assessments have dealt with the response mode issue by measuring writing skills indirectly with multiple choice tests. Support for such indirect measurement derived from reported high correlations between objective tests and direct measures of written production, from the erratic reliabilities accompanying impressionistically scored esssays and from the economic and logistical demands of collecting and scoring writing samples (Braddock, 1963; Godshalk, Swineford & Coffman, 1966; Breland, 1977). More recently, however, demands for writing tests with content, construct and ecological validity have prodded reinstatement of direct, written production tasks, Yet even when assess-

ments collect writing samples, students usually produce only one composition, despite the well documented fluctuation of writing performance from one sample to the next (French, 1962; Braddock, 1963; Diederich, 1974). By necessity, a single sample taps student performance in only one type of discourse and on one topic. This limitation presents a measurement problem since the generic methods of text development in particular forms of discourse differ substantially from one another. Salient structural features of argument, for instance, are issues, reasons, and conclusions, while stories include plot, character, setting, and theme. Exposition involves main idea, supporting detail and logical development; narrative elaborates events in chronological order, and description portrays concrete details in spatial order. Since different purposes set for writing tend to elicit the generic structural elements of the modes of discourse, (Kinneavy, 1971), it seems likely that the schema or frames activated in the writer by writing tasks varying in purpose should differ (Anderson, 1977; Minsky, 1976).

Research on Discourse Mode Effects

Evidence from reading and writing research support the distinctiveness of processes required by varying discourse models. Reading research suggests that different schema are used as students attempt to comprehend narrative and expository text (Meyer, 1975; Graesser, Hauft-Smith, Cohen & Pyles, 1979). As in reading, writers also employ various skills and personal resources to meet the demands intrinsic to

a kind of writing or mode of discourse. For students learning to write, these different discourse modes represent very dissimilar challenges and, futhermore, attempts to compare student writing skills across modes of discourse constitute a real assessment issue. In one of the most frequently cited writing assesment studies, the performance variability in "topic" discussed by Godshalk, et al. cannot separate fluctuations due to the five different discourse modes stimulated by the assignments from that due to the subject matters addressed. Veal and Tillman (1971) reported variability in elementary students' performance on tasks specifying different discourse aims, as did Praeter and Padia (1980).

Moreover, other writing research demonstrates that different writing purposes lead writers to vary structural complexity (Crowhurst & Piche, 1979; Crowhurst, 1980; Perron, 1977) and to represent writing topics quite differently (San Jose, 1973; Perron, 1977). Most importantly for instruction and evaluation, this accumulating body of writing research suggests that different writing purposes require dissimilar writing strategies of unequal difficulty for individual students. Cooper (1979) has cited research indicating that sentence structures shift when discourse mode changes and speculated that a student's planning demands for an essay might change as much as 50%.

The implications for writing assessment of reading and writing studies on discourse mode effects is that the mode of discourse of the writing purpose will make a difference in writing performance. For

example, students might be more skilled at narrative writing tasks requiring chronological development than at expository tasks requiring logical development. Thus the profile of writing competence for a student based on a writing test calling for exposition may differ from the profile of writing competence for that same student on a narrative or persuasive task.

Research on Response Mode Effects

In addition to the question of skill commonality across discourse modes or genres, the question of the response mode or measurement form in which the writing skill should be assessed continues as a hotly debated topic. While many claims are made for the predictive or concurrent validity of indirect, objective writing measures (Coffman, 1971; Breland, 1977), indirect measures simply are not considered by writing researchers to meet the more crucial standards of content or construct validity (Braddock, et al., 1963; Cooper & Odell, 1977). While selected responses elicited by multiple choice tests may provide valuable information, they are, nonetheless, measures of processes required in reading comprehension, not measures of actual production ability. As such, responses to recognition tasks are often considered by learning theorists as, at best, behaviors enroute to constructed responses (Bourne, 1966, Skinner, 1963).

Comparisons of direct and indirect writing measures have yielded moderate correlations between scores from the two response modes. In one of the seminal writing assessment studies, Godshalk, et al. (1966)

reported correlations from .46 to .75 between the sum of five essay scores from high school students and their College Board English Comprehension Test. In an attempt to validate ETS's Test of Standard Written English (TSWE) Breland, Conlan and Rogosa (1976) found correlations of only .42 between that mechanics-oriented test and a 20 minute essay scored on a 4-point scale. In a subsequent study, Breland and Gaynor (1979) reported correlations ranging from .58 to .63 between students' three separate essay scores awarded on a 6-point scale and the TSWE, while the correlation between the sum of the three essays and the TSWE was .76. Similar low to moderate correlations of.43 to .67 were found in a comparison of the American College Testing Program's English Usage Test (also emphasizing sentence-level skills) and students' scores on three essays. Hogan and Mishler's (1980) study of the relationship between third and eighth grade students' Metropolitan Achievement Test scores and one essay yielded correlations of .68 and.65, while the correlations increased to .75 and .81 when a second essay score entered into the calculations.

In general, these studies related essay scores based on norm-referenced holistic ratings to a total multiple choice test score, apparently assuming both sets of measures tapped the same set of writing skills. However, content analyses of the essay rating criteria reveal that, while often vaguely worded, they did reference whole-text features such as thesis, coherence, support and style, as well as sentence-level mechanical conventions. Items on the multiple

choice tests, on the other hand, often emphasized sentence-level mechanics and required few if any text-level discriminations and, obviously, no production responses. At issue then is not simply whether measures correlate statistically, but whether different measures focus on the same text featues of written productions and whether they reflect the same underlying skill constructs.

In a study attempting to compare direct and indirect measures of reasonably parallel text features, Spooner-Smith (1978) used domain-referenced skill specifications to design multiple choice items analogous to essay rating criteria. She found correlations of the multiple choice test total score with a General Impression score of .65 and with the total of analytic ratings of .61. Relationships between analogous features such as Organization and Support, however, were much lower, ranging from .23 to .55. Her findings suggested that when multiple choice scores and essay scores derived from precisely matched definitions of text features, the comparability of scores on these component writing skills might be even lower than those previously reported.

Design Requirements for a Study of Discourse and Response Mode Effects

To compare the information yield and psychometric quality of writing measures involving different discourse and response modes, data are needed that contrast the performance of a group of examinees across equivalently specified skill domains varying only by the modes of measurement. The specific test objectives (skill domains), stimu-

lus dimensions, instructions to examinees, and response criteria/ characteristics need to be matched as closely as possible across discourse and response modes; i.e., each of the measures should present parallel content-valid procedures for assessing the same skill or skills. Data from measures designed to be psychologically parallel can then be examined for evidence of the construct validity and reliability of the discourse and response mode distinctions. A test of writing researchers' contentions that text-level writing skills such as thesis statement, organization and support are best measured by written production tasks would involve comparing multiple choice passage comprehension measures of such subskills as organization, support, and mechanics with ratings of these features in text the student produces. This paper reports such comparisons for score profiles obtained from analytically scored direct assessments of student writing (essay and paragraph length writing samples) and an indirect assessment (multiple choice questions concerning prose passages). Measures were designed to be conceptually parallel by using the domain specifications that guided development of directions/prompts for the writing tasks to construct the passages employed in the multiple choice task. Similarly, the dimensions of writing quality making up the analytical scoring rubric applied to the writing samples determined the specific aspects of the prose passages the multiple choice measure questioned.

The measurement issues addressed in the study concerned the comparability of writing scores obtained from tasks varying in discourse and response mode. The study departed from more conventional methodology in two respects. First, the measures of writing skills in the different response modes were specifically designed to present tasks parallel on all dimensions but the discourse and response mode variables. Second, the study augmented the standard correlational comparisons of the measures with multitrait multimethod (MTMM) factor analyses testing specific hypotheses about sources of variation underlying student's writing score profiles derived from the alternative measures. The analyses examined the convergent and discriminant validity of scores derived from different discourse and response modes, treating the scale scores comprising the writing profiles as "traits," and the discourse or response modes as "methods" (Campbell & Fiske, 1957). Correlations among different operationalizations of the same variable should arise from the influence of a single common factor or trait (e.g., organization). Also, the method of measurement should exert an influence on each variable, so that different variables will covary to a greater degree when measured by a common method than when measured by different methods; this covariation can be thought of as reflecting the operation of a common method factor. This confirmatory factor analytic approach to MTMM validation has been implemented in other empirical studies (Joreskog, 1974; Traub & Fisher, 1977; Werts, Joreskog & Linn, 1972). Traub and Fisher (1977) for example, compared verbal

and quantitative scores derived from fill-in, right/wrong multiple-choice and partial knowledge multiple choice response formats in exactly this fashion.

In this study we asked 1) whether student writing performance profiles are comparable on tasks differing in discourse mode (writing purpose), and 2) whether tasks requiring different response modes (paragraphs, essays, and multiple choice items) provide the same type and quality of information about student writing competence. In the MTMM framework, we examined whether distinctive common factors underlay the corresponding variables from the writing profiles derived from the discourse and response modes variations.

Method

To examine the relationship of writing scores yielded by tasks differing on the two variables, discourse and response mode requirements, high school students received writing tests on three separate occasions. Each student received a multiple choice test and a paragraph writing task, as well as two full length essay assignments. Ratings of the essays and paragraph on an analytic scale and scores on the objective test provided the bases for the comparisons.

Sample

Approximately two hundred eleventh and twelfth grade students from three high schools in a small school district in the Los Angeles area participated in the study. Students were selected who were

attending English or composition classes that were judged by teachers to contain average or above average pupils. Scores from the verbal portion of the Differential Aptitude Test were available for 92 students in the sample; the mean percentile score for this subsample was 63.9 (s.d. = 28.6).

Design

Students within each class were randomly assigned to one of four testing conditons defined by different discourse mode combinations for the essay tasks. In Conditions 1 and 2 (Same Genre), the three constructed response writing tasks (two essays and one paragraph) were in the same discourse mode. Condition 1 students wrote two expository essays and an expository paragraph; Conditon 2 students wrote two narrative essays and a narrative paragraph.

In Conditions 3 and 4, (Different Genre) students wrote one narrative and one expository essay. Condition 3 students wrote an expository essay on Topic A and a narrative essay on Topic B, while Condition 4 students wrote an expository essay on Topic B and a narrative essay on Topic A. Half of the subjects in Conditions 3 and 4 wrote an expository paragraph, while half wrote a narrative paragraph.

Response mode, the second design factor, was a within-subject factor and consisted of the multiple choice test (selected response), the paragraph (short constructed response) and the essay (long constructed response). During the three testing occasions, subjects

received the multiple choice test and paragraph on one occasion, and an essay on each of the other two occasions. The design counterbalanced the order in which students received the tasks.

Measures

The essay and paragraph tasks were constructed in accordance with a set of domain specifications for expository and narrative writing. These specifications included the purpose of the writing assignment, guidelines for appropriate topics, the response criteria by which written products were to be judged, and guidelines for the content and format of the directions for the tasks. The response criteria were chosen to reflect the discourse features of an analytic scoring system developed at UCLA (Pitts, 1978; Spooner-Smith, 1978; Winters, 1978; Quellmalz, 1979). The version of the scoring system used in this study generated five ratings for each written product:

(1) General Impression--A global judgment of writing quality assigned by raters after a quick initial reading of the writing sample.

(2) Focus--The extent to which the subject and main idea of the writing sample were clearly stated or implied.

(3) Organization--The extent to which the main idea was developed according to a discernible method of organization (e.g., clear chronological or logical development).

(4) Support--The extent to which generalizations and assertions were supported by specific, relevant, subordinant statements.

(5) Mechanics--The extent to which the writing sample was free from intrusive sentence-level mechanical errors (e.g., usage, sentence construction, spelling, capitalization and punctuation).

Each essay and paragraph was assigned ratings on these five subscales by one of two pairs of trained reader, the median generalizability coefficient for the two rater pairs were .61 and .83 across topics/occasions and subscales. The three writing samples representing direct measurement (two essays and one paragraph) generated 15 subscale scores, each on a one (low) to four (high) scale. The scores were calculated by averaging the scores assigned by both raters to each written product for each subscale.

The stimulus attributes from the specifications for the writing tasks were used to develop the passages to be read in the multiple choice task. Ten passages were constructed, five expository and five narrative. For each passage, there were three questions, designed to be analogous to text features included in the rating scales--main idea (focus), organization, and support. Main idea questions were referenced to a stated generalization near the beginning or end of the passage. Organization questions required the selection of a new sentence that would best fit at a point in the passage marked by an arrow. Support questions asked which new sentence would best support the main idea of the passage.

Results

Discourse Mode Effects

The first set of analyses compared students' scores according to the discourse mode of the task.

Table 1 presents means and standard deviations of essay ratings for each of the four test conditions.

- - - - - - - - - -

Insert Table 1 here

- - - - - - - - - -

On all five subscales on on total essay scores, narrative ratings were lower than expository ratings. This finding may be due to the differential curricular emphasis given to narrative and expository writing in the high schools, to subjects' lack of knowledge, at a personal experience level, required to deal with the narrative topics, or to raters' tendency to score narratives more stringently.

Table 2 displays the correlations between students' two essay scores on each of the analytic scale subscales. As expected, correlations between essay scores for students writing two essays in the same discourse mode (Same Genre) are higher than those for students in the Different Genre conditions.

- - - - - - - - - -

Insert Table 2 here

- - - - - - - - - -

An examination of the correlations for each subscale across conditions suggests that General Impression and Organization seem to

differentiate most between the different discourse modes (columns 3 and 4 in the table). This finding might also be expected, since General Impression requires a judgment about the global quality of the essay as an example of exposition or narration. Therefore the constellation of essay factors influencing this judgment should be the most comprehensive for each discourse mode and thus the most discriminating. Structurally, exposition and narration differ dramatically in their characteristic use of logical or temporal organizations, respectively. On the Mechanics subscale, correlations across conditions are most comparable, reinforcing the notion that the constellation of syntactic, punctuation, spelling and usage skills may not vary between modes of discourse so much as text-level skills do.

Differences between the correlations pooled within same and different genre conditions reveal that the relationship between student's two essay scores on General Impression, Organization, and the Total is significantly stronger when students write in the same genre than when they write in different genre.

Only between-group comparisons were conducted between discourse modes for the paragraph data since each student wrote only one paragraph. Table 3 presents the results of these comparisons. The same analytic scales for narration and exposition used for rating the essays were also used for rating the paragraphs.

- - - - - - - - - - - - - -

Insert Table 3 about here

- - - - - - - - - - - - - -

Subscale scores ranged from 1-4, total scores from 1-20. Ratings of narrative paragraphs were generally lower than ratings of expository paragraphs. Ratings of expository paragraphs differed significantly from narrative paragraphs on the General Impression, Focus, and Organization subscales and on the Total scores. Consonant with essay data, Mechanics and Support were not as influenced by the different discourse tasks.

Multiple choice test comparisons of interest were the scores each individual received on the narratve and expository sections of the exam. Table 4 presents the means and standard deviations.

- - - - - - - - - - - - -

Insert Table 4 about here

- - - - - - - - - - - - -

On this reading comprehension test of items measuring recognition of writing-related skills, students were able to answer Focus/Main Idea and Support questions similarly well for both expository and narrative passages. On Organization questions, however, students had more difficulty in general (73% overall average), particularly with narrative organization (66%). Correlations between individual multiple choice subscales across genre range from .47 to .49, with the between-genre total score correlation equalling .65; these figures suggest little differential sensitivity to genre for the subscales, comparing most closely to the within-genre correlations reported in Table 2.

In the aggregate, the preceding analyses contrasting students' performance on writing tasks differing in discourse mode suggest that: 1) students' writing skills vary in the different discourse modes, and 2) discourse mode score variability seems to be differentially distributed across writing subskills. These results occurred in separate analyses of students' essays, paragraphs and multi-choice scores.

To test the effect of discourse mode on subskills, the data were then subjected to multi-trait multi-method (MTMM) analyses using confirmatory factor analysis techniques (Joreskog, 1974). Traits were defined as the writing subskills; methods were defined as exposition and narrative. The analyses required within-subject measures of trait and discourse mode, therefore data from 94 subjects in test conditions 3 and 4 were used to examine the effects of discourse mode. In these conditions students wrote an expository and narrative essay and answered multiple choice questions about expository and narrative passages. Since students wrote both expository and narrative paragraphs, paragraph scores were not included in the analyses.

To examine the factor structure of discourse modes in essay performance, the five analytic subscales of General Impression, Focus, Organization, Support and Mechanics formed the trait dimensions. The five trait indicators for each genre were formed by averaging scores over raters and standardizing across topics. Thus ten scores were constructed, two each (expository and narrative) for each of the five subscales.

To examine the factor structure of discourse modes across response modes, a second set of analyses used just the three subscales (traits) common to the essay and multiple choice tests, Focus, Organization, and Support. For the multiple choice test, number correct scores were formed within discourse mode for each of the three multiple choice subscales. This second set of analyses employed six scores, two each (expository and narrative) for Focus, Organization and Support.

Maximum likelihood estimates of the parameters of the MTMM confirmatory factor analysis models were obtained from the LISREL computer program for the analysis of covariance structures (Joreskog, 1973, 1977; Joreskog & Sorbom, 1978). The LISREL program allows the analyst to treat model parameters (e.g., factor loadings or factor intercorrelations) in one of three ways: (a) as free parameters to be estimated by the program; (b) as fixed parameters specified in advance to equal some fixed number (usually zero); or (c) as constrained parameters to be estimated by the program subject to the constraint that they equal other estimated parameters. In addition, the program computes standard errors for all free and constrained parameters, as well as an overall chi square test of the model's fit to the data. All model equations (in LISREL notation) are of the form:

$$\underset{\sim}{Y} = \underset{\sim}{\Lambda}_Y \underset{\sim}{\zeta} + \underset{\sim}{\varepsilon} \qquad (1)$$

$$\underset{\sim}{\Sigma}_Y = \underset{\sim}{\Lambda}_Y \underset{\sim}{\Psi} \underset{\sim}{\Lambda}_Y' + \underset{\sim}{\theta}_\varepsilon \qquad (2)$$

Each observed score in Y depends on the latent variables $\underset{\sim}{\xi}$ and $\underset{\sim}{\varepsilon}$, common factors and measurement errors, respectively. Equation (2) shows the hypothesized structure underlying the covariance matrix of the Y's; it consists of a matrix of factor loadings $\underset{\sim}{\Psi}$ (hereafter Psi), and a (usually) diagonal matrix of error variances. Interest in the analyses to follow focuses on the contents of Lambda and Psi.

The first set of MTMM analysis examined the influence of subscale content (i.e., Focus, Support, etc.) and essay genre (Expository and Narrative) on the scores for students writing essays in different discourse modes. Eight scores are entered into the analysis, the measure of Focus, Organization, Support, and Mechanics for each essay.[1] Correlations among the scores are shown in Table 5, and a path diagram of the initial factor analysis model in Figure 1, Panel A.

Insert Table 5 and Figure 1 here

In Panel B of the figure is shown the path diagram for the final model for these eight writing scores. In moving from the initial to the final model a number of intermediate analyses were run, each refining the factor structure until a completely adequate fit to the data was obtained. Parameter estimates for the final model are shown in Table 6.

Insert Table 6 about here

From the path diagram in Panel B, it can be seen that only two of the original four trait factors emerge in the final model, while "method" factors corresponding to discourse modes remain in the model. The one assymetry between discourse modes is the behavior of the Organization subscales: the measure of organization from the expository essay apears to be more or less fully saturated with genre variation, which, when removed, leaves an insignificant remainder of reliable variation to be shared with the other measures of coherence. The table shows the matrix of estimated factor loadings. All of the estimated parameters in the model are statistically different from zero. Except for Mechanics, all of the subscales contain relatively larger components of genre-related variation than "trait" variation (for both Support measures and Expository Organization, genre completely dominate). Variation in student Mechanics, on the other hand appears comparatively robust to the influence of genre. The factor intercorrelation matrix (Psi) is not included in the Table since all correlations are zero. While factor intercorrelations were permitted in the initial model, the factors which remain in the final solution are orthogonal.

In fitting the MTMM models to the writing score data, we attempted, where possible, to constrain appropriate pairs of factor loadings to equal one another. That is, it was usually of interest to check whether, for example, the loadings of the Focus variable derived from an Expository essay equalled those of the Focus variable from a Narra-

tive essay. We see in the table that such equality constraints were incompatable with the subscale variables' relations to the genre factors (the first two columns of the Lambda matrix), while they are compatable for subscale relations to the trait factors (where present). Finally, the table reports a non-significant goodness-of-fit statistic (chi square on 17 df = 18.678; p = .347), indicating this model cannot be rejected.

The overall picture of genre effects on writing presented by this analysis is that, except for relatively lower level skills (mechanics), the discourse mode students are required to write in is a strong influence on their performance. The different subskills included in the scoring rubric seem definitely to interact with discourse mode and, at the same time, to varying degrees are independent sources of variation in student writing performance.

Additional information was collected during the study which bears on possible genre effects on writing. For three of the four subscales included in the previous analysis, multiple choice passage comprehension subtest scores were available for Expository and Narrative passages. We take up these six variables in the next set of analyses. With only six observed scores, it is not possible to specify a full MTMM model as in the previous analysis. Furthermore, one of our hypotheses concerns the adequacy of multiple choice measurements for obtaining distinctive information about writing subskills. With these concerns in mind, we specified three alternative factor models for the

multiple choice scores; these are displayed in Figure 2 as Panels A, B, and C.

Insert Figure 2 about here

Panel A shows a one-factor model, corresponding to the hypothesis that all of the multiple choice scores measure the same thing,e.g., general comprehension. Panel B shows a two-factor model, testing the possibility that the genre in which each of the passages is written determines how the resulting scores covary. Lastly, Panel C shows a factor model with three traits, one for each subscale. Although these models cannot be strictly compared in a statistical sense (i.e., they are not hierarchically related to one another), gross differences in goodness-of-fit can guide the selection of an appropriate final model. Correlations among the multiple choice variables are shown in Table 7. A summary of the best fitting models is presented in Table 8.

Insert Tables 7 and 8 here

The two-factor model was found to fit less well than either the one or three-factor models ($p = 0.000$), suggesting that, contrary to the picture of genre effects observed for actual writing tasks, passage genre was not a source of discriminant validity for multiple choice scores. The one- and three-factor final models have identical

degrees of freedom and goodness-of-fit statistics, indicating that these are statistically equivalent models. In the one-factor model, this degree of fit was obtained by allowing residual covariation among the pairs of measurement errors associated with each trait--what is common among the scores is captured by the single common factor and what is unique to each trait by the residual covariances. In the three-factor model, the common variance and residual covariance are combined into three correlated factors. In a sense, then, we might equally well interpret the multiple choice scores as measuring three correlated factors. We are led to prefer the latter interpretation by the fact that the standard errors for the estimates of factor intercorrelations in the three-factor model are sufficiently small as to make it highly improbable that each of these correlations would actually be equal to unity, the situation that would need to obtain if a one factor model were appropriate. We should note that an additional two-factor model was tried, combining the focus and organization multiple choice scores into a single Coherence factor, that produced a considerably poorer fit than the one- and three-factor models shown in Table 8.

Response Mode Effects

The second measurement issue addressed by the study was whether tasks requiring different response modes (direct production modes: essays, paragraphs; indirect selection modes: multiple choice) provide the same type and quality of data about student writing abilities.

An MTMM factor analysis was performed to examine this issue. The "method" variables for response mode were constructed according to procedures analogous to those used to construct the discourse mode variables. Each of the five subscale scores provided by the essay and paragraph ratings was averaged over raters, then standardized within topic and discourse mode. Standardizations were employed to reduce possible interactions between response mode and genre and/or topic. Number correct scores were formed within genre for each of the three multiple choice subscales, standardized within genre, summed across genre, and then restandardized to produce scores scaled in a manner comparable to those derived from the writing samples. No measures of Mechanics were included in the multiple choice task. Complete data were available for 148 of the students. In sum, 15 scores were constructed for analysis, three measures of Mechanics (two essay and one paragraph), and four each of focus, Organization and Support (two essay, one paragraph and one multiple choice). Table 9 presents the correlations matrix for the three response modes and four subscales.

- - - - - - - - - -

Insert Table 9 here

- - - - - - - - - -

The MTMM analyses began by considering the data for the two essays each student wrote, i.e., "Essay 1" and "Essay 2" methods only; the first analysis included the eight scores defined for these two conditions, two measures of focus (fe_1, fe_2), Organization (oe_1, oe_2),

Support (se_1, se_2), and Mechanics (me_1, me_2). The model specified for these variables includes the three relatively distinct trait/subscale content factors emerging from the discourse mode analyses and two "method" factors, one for each essay. In keeping with the results from the final factor analysis of the discourse mode data, the focus and organization subscales are combined into a single trait/content factor, Coherence. Figure 3 displays a path diagram for this model and Table 10 presents the LISREL estimates of the free and constrained parameters for the model.

- - - - - - - - - - - - - - - - - -

Insert Figure 3 and Table 10 here

- - - - - - - - - - - - - - - - - -

As in the model for the discourse mode MTMM analysis, trait factor intercorrelations are unconstrained in this model, while the method factors are constrained to be uncorrelated with each other and with the subscale/trait factors. This latter restriction reflects our hypothesis that the method factors act as independent additive components influencing observed scores. In addition, trait factor loadings for pairs of subscale measures have equality constraints placed on them. These constraints test whether subscale scores from different essays exhibit the same degree of relationship to the traitfactor they measure.

The model as a whole cannot be rejected; the non-significant chi square ($p \geq .202$) suggests that the model provides an adequate account

for the observed correlations among essay variables. Loadings of the essay variables on their respective trait factors are all substantial, ranging from a low of .472 for Organization to a high of .768 for Mechanics. Method factor loadings for the Organization subscale are relatively high while those for the Support subscale are moderate. Both Focus and Mechanics method loadings are low. In terms of the model's decomposition of observed scores into trait and method components, then, Focus and Mechanics show stronger dependence on trait than method factors and Support is about equally related to both sources. Organization, however, is more strongly influenced by method variance, reflecting, in part, the genre differences for this subscale found in the earlier analyses. Turning to the psi matrix, estimates of the relations among the trait factors are moderate (below .80), ranging from a low of .626 for the correlation between Mechanics and Support to a high of .799 between Coherence and Support. Mechanics appears to be the most independent of the three content factors. The next MTMM model adds data from the paragraph task and expands to 12 the number of variables in the analysis. Each of the four new subscale scores is specified to load on the same trait factor as did the analogous essay subscale in the previous analysis (no constraints are placed on these additional loadings); and a third method factor, Paragraph, is included to account for covariation specific to this mode of responding. Figure 4 displays the path diagram for this model and Table 11 presents the results of the LISREL estimation of the parameters of this model.

- - - - - - - - - - -

Insert Figure 4 here

- - - - - - - - - - -

- - - - - - - - - - -

Insert Table 11 here

- - - - - - - - - - -

The model provides an adequate overall fit to the observed intercorrelations among the essay and paragraph variables (chi square on 43 df = 51.533, $p \leq .175$). This is consistent with the hypothesis that the scores generated by application of the rating system to paragraph length writing samples can be interpreted as measuring the same underlying content as the scores derived from full length essays. Inspection of the lambda matrix shows that the loadings for paragraph subscale scores on their associated trait factors are of substantial magnitude in each case, and that the loadings on the paragraph factor follow the same general pattern as for the two essay method factors. With one exception, the paragraph variables appear to relate to trait factors less strongly than do the essay scores. The one exception is an interesting one: "sp" provides a clearer definition of the Support factor than either of the support measures derived from essays. This would seem to suggest that the rater's task of judging the use of support is carried out more distinctly in the context of single paragraphs than it is in longer writing samples.

As for the essay-only model, the trait intercorrelations for the paragraph and essay model are moderate to high, indicating considerable interdependence among the subscales. Again, Mechanics exhibits lower levels of relationship to the other subscales.

Comparison of these two response mode models reveals two main differences. First, there is some instability in the size of the essay variables' loadings on the associated trait factors as we move from the first to the second model. This leads to the interpretation that the factors composed of both essay and paragraph variables do not measure precisely the same content as factors composed of essay variables only. Second, estimates of the trait intercorrelations in the second model are greater than their counterparts in the first model. Thus, although the inclusion of paragraph scores may have broadened the content of the trait factors, it seems also to have diminished their distinctiveness.

The third MTMM analysis builds on the previous two by adding the three scores derived from the multiple choice items administered to the students in the study. Recall that only items analogous to the Focus, Organization and Support subscales were included in the multiple choie test. This model differs from the previous one, then, by the specification of trait loadings for these three subscales, and the addition of a multiple choice method factor. Figure 5 displays the path diagram and Table 12, the LISREL estimates of the model parameters, for the analysis of data from all three response modes.

- - - - - - - - - - - - - - - - - -

Insert Figure 5 and Table 12 here

- - - - - - - - - - - - - - - - - -

As in the first two analyses, the model provides a reasonably good fit to the data (chi square on 76 df = 84.952, p = .226), implying that the same 3-trait structure is not violated by the inclusion of the multiple choice scores. By and large, however, estimates of the essay trait factor loadings have dropped in value in comparison with the corresponding estimates from previous models. Also, the trait factor intercorrelations have increased for Coherence and Mechanics, indicating that the subscale content factors have drifted closer together as a result of adding the multiple choice variables. Thus, while the multiple choice scores apparently share some content with the constructed response variables to which they are purportedly analogous, they also seem to possess a higher degree of "latent collinearity" (Yates, 1979) in the trait factor space. Whether this situation arises because the multiple choice variables are related to writing ability in some non-specific fashion, or because all of the variables, but especially the multiple choice scores, share a common dependence on general verbal ability, cannot be disentangled without additional analyses, including tests marking general ability factors. In any event, it is reasonable to interpret the increased interdependence among trait factors as an indication that the multiple choice scores possess generally lower validity as indices of distinctive

components of writing ability than do measures based on actual writing samples.

The final model examines the relationship of the paragraph and multiple choice variables to the set of trait factors defined solely on the basis of the essay variables. Generally speaking, both the contribution of the essay scores to the definition of the subscale content factors and the degree of independence of these factors from one another were reduced as data from the alternative response modes were added to the analysis. In the final model the trait factor structure obtained when only essay scores were included in the analysis (cf., Figure 3 and Table 10) as a criterion definition of the content underlying the subscales, treating the earlier trait factors as "unmeasured" criterion variables against which to compare the scores from the other two response modes. This can be accompllished in LISREL by modifying the specification for the third model (see Table 12) in two ways. First, instead of estimating trait loadings for essay variables, new specifications fix their values to equal those estimated from the essay data alone. Second, we place a similar constraint on the trait factor intercorrelations in Psi, by fixing their values at those obtained in the essay-only solution. These two sets of restrictions will ensure that the essay-only trait factors will be reproduced exactly, and the standing of the paragraph and multiple choice variables can be evaluated vis-a-vis the essay criterion trait structure. The LISREL estimates of the free parameters of the final model are contained in Table 13.

- - - - - - - - - - -

Insert Table 13 here

- - - - - - - - - - -

The only parameter estimates of direct interest in Table 13 are the trait factor loadings for the paragraph and multiple choice variables. The data indicate near uniform reduction in their magnitude in comparison to the estimates obtained when trait factors are fit to data from all three response modes. This shift does not reduce the overall model fit (chi square with 83 df = 95.547, p = .164). In all but two instances, paragraph and multiple choice trait factor loadings are lower than the corresponding loadings for the essay variables. Both exceptions are recurrences of the findings from the second and third models that the measure of Support derived from a paragraph length writing sample outperforms the Support measures based on full length essays and the Organization score in multiple choice is slightly more distinct than the Organization measures on essays respectively. Support, as measured by multiple choice items, seems to reflect relatively little of what is measured in actual writing samples; while multiple choice measures of Focus and Organization seem to convey a roughly comparable amount of information about subscale content to that contained in a single paragraph.

Summary and Conclusions

The purpose of the study was to examine the comparability of writing competency profiles derived from test tasks differing in

discourse and response mode. Theory and research in the fields of learning, instruction and rhetoric have fueled contentions that the knowledge structures and processing strategies activated by different writing aims and modes of responding are quite distinct. We were attempting to demonstrate the robustness of these claims from a measurement perspective.

In practice, many current writing assessment programs fail to consider the validity of test data that does not distinguish between the demands of types of writing tasks and between the requirements of production and selection. At heart, the issue is one of construct validity, do these alternative task and processing variables measure the same thing? Our results indicate that the answer is "no."

In this study the results of correlational, parametric and multi-trait multi-method analyses indicate that levels of performance vary on tasks presenting different writing purposes. These data cast doubt on the assumption that "a good writer is a good writer" regardless of the assignment. The implication is that writing for different aims draws on different skill constructs which must therefore be measured separately to avoid erroneous, invalid interpretations of performance. The findings suggest that generalizations about student writing competence must reference the particular discourse domain rather than the general domain of writing.

The study also investigated the distinctiveness of information about writing compoetence provided by direct and indirect measurement

techniques. Here again, the issue is one of validity--do both response modes measure the same skill construct? Our results suggest this question cannot be answered unequivocally. Tests of response mode effects within MTMM confirmatory factor-analytic framework show method variance to be present in varying degrees. Organization is particularly influenced by the response requirements of the task (organization was also most sensitive to genre). The patterning of subscales' saturation with method variance is consistent within constructed response tasks. Writing variables measured in a multiple choice format, however, follow a different pattern of method variation and, while no more strongly saturated with systematic error, show generally lower levels of communality (i.e., more random measurement error).

While models can be fitted to the data from all three response modes that confirm the subscales' content, the degree of independence of the resulting subscale factors appears to be affected by which response modes are included in the analysis. The most differentiated subscale factor structure is obtained by including only essay variables in the analysis; interdependence among the subscale factors increases with the addition of both paragraph and multiple choice measures. Thus, the effect of shortening the assessment task for the examinee through examination of just paragraphs or of changing the form of the response (multiple choice tasks) does not simply increase the measurement error. The savings in testing time are obtained also

at the cost of clarity and distinctiveness in the information about each of the subscales. When the subscale content factors are located in the variable space so as to maximize their relationship to scores derived from the essay response mode, most other subscale-response mode combinations provide weaker substantive information. The main exception is the measure of Support based on paragraph-length writing samples which seems to be superior to the corresponding essay variables in its ability to capture subscale content. It may be that the use of support is less equivocally evaluated in the context of a single paragraph than in an essay containing multiple paragraphs, each of which may suggest a different view of the examinee's ability to provide supporting detail.

The MTMM analyses also provided information about the validity of the rating scales. The MTMM analyses suggested, first, that repeated applications of the scoring rubric developed in our reseach to writing samples in fact produce measures that tap the same underlying content. Thus, given multiple measures of each subscale, it is possible to fit a factor analysis model that confirms their hypothesized content. Second, it was found that factors reflecting the content of the writing subscales are strongly intercorrelated, and this interdependence appears to be present no matter what response mode subjects are assessed in. When the global judgment for General Impression was removed and Focus and Organization combined into a Coherence subscale, scale intercorrelations became more moderate and distinct. Since

techniques for producing writing that is coherent, supported and mechanically correct are often taught separately, further examination of the value of rating writing accoring to separate component features should consider both their diagnostic utility and component distinctiveness.

Finally, the study exemplified the contribution MTMM analyses can make to validity studies. The technique may provide more sensitive, precise statistical indices of hypothesized competencies underlying test performance.

In summary, the study highlights the importance of precision in designing, analyzing and reporting writing assessment data. It may be that the techniques developed for specifying domain-referenced skill boundaries can provide a reasonable framework or focusing attention, discussion, assessment and instruction on clearly bounded classes of writing performance.

Footnote

[1]Note that we have excluded from these and subsequent analyses the General Impression ratings. In preliminary factor analyses including these scores, we found them to be inseparable from Focus and Organization ratings, suggesting that raters' general impressions contain little or no additional information.

References

Anderson, R. C. The notion of schemata and the educational enterprise. In Anderson, R. C., Spiro, R. J., & Montague, W. E. (Eds.), Schooling and the acquisition of knowledge. Hillsdale, NJ: Lawrence Earlbaum Associates, 1977.

Bourne, L. J. Human conceptual behavior. Boston: Allyn and Bacon, 1966.

Braddock, R., Lloyd-Jones, R., & Schoer, L. Research in written composition. Champaign, IL: National Council of Teachers of English, 1963.

Breland, H. M. A study of college English placement and the test of standard written English. Princeton, NJ: Educational Testing Service, January, 1977.

Breland, H. M., Conlan, G. C., & Ragosa, D. A preliminary study of the Test of Standard Written English. Princeton, NJ: Educational Testing Service, 1976.

Breland, H. M., & Gaynor, J. L. A comparison of direct and indirect assessments of writing skill. Journal of Educational Measurement, 1979, 16, 119-128.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validity by the multitrait-multimethod matrix. Psychological Bulletin, 1979, 56, 81-105.

Coffman, W. E. Essay exams. In R. L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D. C: American Council on Education, 1971.

Cooper, C. R. Current studies of writing achievement and writing competence. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1979.

Cooper, C., & Odell, L. (Eds.) Evaluating writing: Describing, measuring, judging. Buffalo, NY: State University of New York at Buffalo, 1977.

Crowhurst, M. Syntactic complexity in narration and argument at three grade levels. Canadian Journal of Education, 1980.

Diederich, P. B. Measuring growth in English. Champaign, IL: National Council of Teachers of English, 1974.

French, J. W. Schools of thought in judging excellence of English themes. 1961 Proceedings of Invitational Conference on Testing Problems, Educational Testing Service, Princeton, NJ, 1962.

Godshalk, F. I., Swineford, F., & Coffman, W. E. The measurement of writing ability. New York: College Entrance Examination Board, 1966.

Graesser, A. C., Hauft-Smith, K., Cohen, A. D., & Pyles, D. Familiarity and test genre on retention of prose. Journal of Experimental Education, 1979, 48, 281-290.

Hogan, T. P., & Mishler, C. Relationships between essay tests and objective tests of language skills for elementary school students. Journal of Educational Measurement, 1980, 17, 219-227.

Jöreskog, K. G. Analyzing psychological data by structural analysis of covariance matrices. In R. C. Atkinson, D. H. Krantz, & P. D. Suppes (Eds.), Contemporary Development in Mathematical Psychology, Vol. II. San Francisco: W. H. Freeman & Co., 1974, 1-56.

Kinneavy, J. L. A theory of discourse. The Aims of Discourse. Englewood Cliffs, NJ: Prentice Hall, Inc., 1971.

Meyer, B. F. The organization of prose and its effects on memory. North Holland Studies in Theoretical Poetics (Vol. I). Amsterdam: North Holland Publishing Company, 1975.

Minsky, M. A framework for representing knowledge. In P. H. Winston (Ed.), The psychology of computer vision. New York: McGraw-Hill, 1975.

Perron, J. D. Written syntactic complexity in modes of discourse. Paper presented at the annual meeting of the American Educational Research Association, New York, April 1977.

Pitts, M. The relationship of classroom instructional characteristics and writing in the descriptive/narrative mode. Report to the National Institute of Education, November, 1978. (Grant No. OB-NIE-G-78-0213 to the UCLA Center for the Study of Evaluation.)

Praeter, D., & Padia, W. Effects of modes of discourse in writing performance in grades four and six. Paper presented at the annual meeting of the American Educational Research Association, Boston, 1980.

Quellmalz, E. Interim Report. Defining writing domains: Effects of discourse and response mode. Center for the Study of Evaluation, 1979.

San Jose, C. P. M. Grammatical structures in four modes of writing at fourth grade level. Unpublished doctoral dissertation, Syracuse University, 1972. Dissertation Abstracts International, 1973, 33, 5411-A.

Skinner, B. F. Teaching machines and programmed learning. New York: Appleton Century Crofts, 1963.

Smith, L. S. Investigation of writing assessment strategies. Report to the National Institute of Education, November, 1978. (Grant No. OB-NIE-G-78-0213 to the UCLA Center for the Study of Evaluation.)

Traub, R. E., & Fisher, C. W. On the equivalence of constructed-response and multiple-choice tests. Applied Psychological Measurement, 1977, 1(3), 355-369.

Veal, L. R., & Tillman, M. Mode of discourse variation in the evaluation of children's writing. Research in the teaching of English, 1971, 5, 37-45.

Werts, C. E., Jöreskog, K. G., & Linn, R. L. A multitrait-multimethod model for studying growth. Educational and Psychological Measurement, 1972, 32, 655-678.

Winters, L. The effects of differing response criteria on the assessment of writing competence. Report to the National Institute of Education, November, 1978. (Grant No. OB-NIE-G-78-0213 to the UCLA Center for the Study of Evaluation.)

Yates, A. Distinguishing trait from method variance in fitting the invariant common factor model to observed intercorrelations among personality rating scales. Paper presented at the annual meeting of the Society of Multivariate Experimental Psychology, Los Angeles, CA, 1979.

Table 1

Means and Standard Deviations of Essay Scores

Test Condition		Same Genre				Different Genre			
		1		2		3		4	
Topic		A	B	A	B	A	B	A	B
General Impression	$\bar{X}$	2.20	2.01	1.27	.88	1.43	1.89	.94	2.09
	sd	.60	.70	1.44	1.09	1.89	.66	.88	.54
Focus	$\bar{X}$	2.45	2.30	2.21	2.21	2.28	2.19	2.26	2.33
	sd	.69	.64	.68	.64	.59	.69	.51	.53
Organization	$\bar{X}$	2.16	1.98	1.88	1.62	1.88	1.95	1.52	2.08
	sd	.72	.70	.98	.71	.85	.70	.56	.56
Support	$\bar{X}$	2.34	2.38	2.42	2.19	2.36	2.28	2.10	2.26
	sd	.64	.54	.81	.66	.70	.65	.54	.53
Mechanics	$\bar{X}$	2.35	2.34	2.56	2.42	2.28	2.03	2.35	2.42
	sd	.64	.68	.69	.67	.80	.81	.56	.54
Total	$\bar{X}$	11.50	10.90	10.35	9.32	10.23	10.31	9.18	11.19
	sd	2.78	2.67	3.84	2.89	3.36	3.02	2.02	2.05
n =		40	40	39	39	40	40	54	54

Table 2

Correlation Between Students' Two Essays

	Same Genre		Different Genre		z difference[1]
Condition	1	2	3	4	1&2 vs. 3&4
General Impression	.56	.39	.33	-.08	2.52*
Focus	.43	.68	.41	.37	1.50
Organization	.42	.42	.20	-.10	2.55*
Support	.27	.28	.07	.38	.31
Mechanics	.58	.68	.63	.50	.65
Total	.60	.55	.41	.22	2.10*
n =	40	39	40	54	

[1]Difference between correlations for conditions 1 and 2 vs. conditions 3 and 4 after r-to-z transformation.
*$p \geq .05$

Table 3

Difference Between Scores on Expository and Narrative Paragraphs

		Expository	Narrative	t diff.
General Impression	$\bar{X}$	1.93	1.05	6.74***
	s.d.	.71	1.06	
Focus	$\bar{X}$	2.34	2.09	2.93**
	s.d.	.54	.63	
Organization	$\bar{X}$	1.95	1.70	2.41**
	s.d.	.62	.78	
Support	$\bar{X}$	1.94	2.02	.87
	s.d.	.67	.71	
Mechanics	$\bar{X}$	2.35	2.29	.65
	s.d.	.65	.74	
Total	$\bar{X}$	10.50	9.15	3.37**
		n = 111	n = 89	

**$p \leq .01$

***$p \leq .001$

Table 4

Means and Standard Deviations of Multiple-Choice Test Scores

		Expository	Narrative	Total
Focus	$\bar{X}$	4.61 (92%)	4.50 (90%)	9.13 (91%)
	s.d.	.77	.89	1.38
Organization	$\bar{X}$	4.05 (81%)	3.31 (66%)	7.39 (73%)
	s.d.	1.03	1.13	2.03
Support	$\bar{X}$	4.56 (91%)	4.41 (88%)	8.97 (90%)
	s.d.	.77	.98	1.51
Total	$\bar{X}$	13.23 (88%)	12.23 (82%)	25.33 (84%)
	s.d.	2.02	2.43	4.26

n = 241

Table 5

Correlations Among Essay Writing Scores
From Different Discourse Modes

	FON	ORN	SUN	MEN	FOE	ORE	SUE	MEE
FON	1.00							
ORN	.690	1.00						
SUN	.474	.600	1.00					
MEN	.351	.286	.322	1.00				
FOE	.327	.330	.060	.222	1.00			
ORE	.108	.007	.116	.125	.456	1.00		
SUE	.066	-.016	.087	.297	.461	.478	1.00	
MEE	.083	.015	.109	.545	.386	.354	.410	1.00

N = Narrative
E = Expository
F = Focus
O = Organization
S = Support
M = Mechanics

Table 6

LISREL Estimates for Final MTMM Model for Essay Writing Scores from Different Discourse Modes

Lambda Matrix[1]

	Coherence	Mechanics	Narrative	Expository
	.528[2]	0.0	.562	0.0
FOE	.528[2]	0.0	0.0	.717
ORN	.607	0.0	.696	0.0
ORE	0.0*	0.0	0.0	.682
SUN	0.0	0.0	.866	0.0
SUE	0.0	0.0	0.0	.687
MEN	0.0	.657[3]	.346	0.0
MEE	0.0	.657[3]	0.0	.437

[1]All estimated coefficients exceed twice their standard errors.

[2,3]Pairs of loadings constrained equal.

*All zero coefficients fixed a priori.

Table 8

LISREL Estimates for One- and Three-factor Models for Multiple Choice Variables from Different Discourse Modes

1-Factor	Lambda	Residual Covariance
FMN	.638	FMN x FME .068
FME	.638	
OMN	.578	OMN x OME .155
OME	.578	
SMN	.631	SMN x SME .095
SME	.631	λ^2_9 = 21.167 ($p \leq .012$)

3-Factor	Lambda		
	Focus	Organization	Support
FMN	.689	0.0	0.0
FME	.689	0.0	0.0
OMN	0.0	.699	0.0
OME	0.0	.699	0.0
SMN	0.0	0.0	.703
SME	0.0	0.0	.703

	Psi		
	Focus	Organization	Support
Focus	1.0		
Organization	.766	1.0	
Support	.832	.743	1.0

λ^2_9 = 21.167 ($p \leq .012$)

Table 9

Correlation Matrix of Scores Containing Response Mode and Subscale Trait Effects

	FOE_1	ORE_1	SUE_1	MEE_1	FOE_2	ORE_2	SUE_2	MEE_2	FOP	ORP	SUP	MEP	FOMC	ORMC	SUMC
FOE_1	1.000														
ORE_1	.441	1.000													
SUE_1	.438	.566	1.000												
MEE_1	.423	.437	.396	1.000											
FOE_2	.424	.211	.247	.234	1.000										
ORE_2	.345	.252	.230	.264	.608	1.000									
SUE_2	.233	.247	.314	.266	.475	.550	1.000								
MEE_2	.365	.276	.299	.589	.512	.419	.361	1.000							
FOP	.231	.057	.252	.305	.270	.192	.186	.353	1.000						
ORP	.196	.176	.223	.233	.279	.228	.170	.297	.559	1.000					
SUP	.324	.269	.353	.297	.375	.322	.294	.428	.520	.579	1.000				
MEP	.345	.311	.314	.603	.352	.340	.272	.555	.433	.429	.330	1.000			
FOMC	.256	.226	.149	.222	.379	.353	.215	.373	.362	.286	.344	.311	1.000		
ORMC	.229	.142	.181	.285	.360	.307	.256	.344	.304	.339	.328	.352	.474	1.000	
SUMC	.239	.312	.197	.317	.264	.253	.172	.354	.259	.291	.245	.360	.419	.437	1.000

Table 10

LISREL Estimates for Response Mode and Trait Effects in Essay Data

Lambda Matrix*

	Coherence	Support	Mechanics	$Essay_1$	$Essay_2$
Fe_1	.633	0	0	.255	0
Fe_2	.633	0	0	0	.485
Oe_1	.472	0	0	.671	0
Oe_2	.472	0	0	0	.639
Se_1	0	.535	0	.534	0
Se_2	0	.535	0	0	.597
Me_1	0	0	.768	.289	0
Me_2	0	0	.768	0	.254

PSI	C	S	M
C	1.0		
S	.799	1.0	
M	.706	.626	1.0

$\lambda^2_{13} = 16.940$ ($p \leq .202$)

*All estimated coefficients exceed twice their standard errors.

Table 11

LISREL Estimates for Response Mode and Trait Effects in Essay and Paragraph Data*

Lambda Matrix

	C	S	M	E_1	E_2	Paragraph
Fe_1	.581	0	0	.282	0	0
Fe_2	/581	0	0	0	.495	0
Fp	.485	0	0	0	0	.473
Oe_1	.468	0	0	.683	0	0
Oe_2	.468	0	0	0	.650	0
Op	.417	0	0	0	0	.780
Se_1	0	.522	0	.502	0	0
Se_2	0	.522	0	0	.461	0
Sp	0	.639	0	0	0	.421
Me_1	0	0	.773	.250	0	0
Me_2	0	0	.773	0	.192	0
Mp	0	0	.728	0	0	.207

PSI	C	S	M
C	1.0		
S	.937	1.0	
M	.809	.684	1.0

$\hat{\lambda}^2_{43} = 51.333$ ($p \leq .175$)

*All estimated coefficients exceed twice their standard errors.

Table 12

LISREL Estimates for Response Mode and Trait Effects in Essay, Paragraph and Multiple Choice Data

Lambda Matrix

	C	S	M	E_1	E_2	P	Multiple Choice
Fe_1	.585	0	0	.323	0	0	0
Fe_2	.585	0	0	0	.472	0	0
Fp	.514	0	0	0	0	.444	0
Fmc	.546	0	0	0	0	0	.399
Oe_1	.483	0	0	.672	0	0	0
Oe_2	.483	0	0	0	.619	0	0
Op	.470	0	0	0	0	.737	0
Omc	.528	0	0	0	0	0	.462
Se_1	0	.469	0	.546	0	0	0
Se_2	0	.469	0	0	.480	0	0
Sp	0	.611	0	0	0	.398	0
Smc	0	.483	0	0	0	0	.397
Me_1	0	0	.772	.277	0	0	0
Me_2	0	0	.772	0	.183	0	0
Mp	0	0	.747	0	0	.182	0

PSI	C	S	M
C	1.0		
S	.978	1.0	
M	.786	.786	.786

$\lambda^2_{76} = 84.962$ ($p \leq .226$)

*All estimated coefficients exceed twice their standard errors.

Table 13

LISREL Estimates for Response Mode and Trait Effects Using Essay Trait Factors as Criteria*

Lambda Matrix

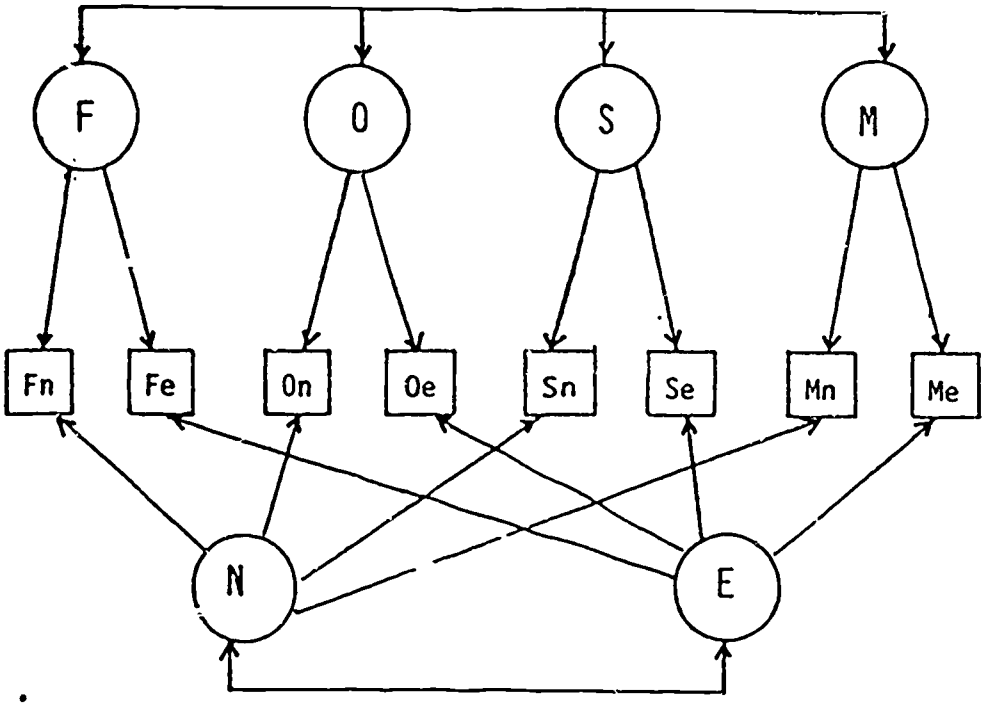
	C	S	M	E_1	E_2	P	Mc
Fe_1	.633[1]	0	0	.333	0	0	0
Fe_2	.633[1]	0	0	0	.436	0	0
Fp	.463	0	0	0	0	.457	0
Fmc	.526	0	0	0	0	0	.407
Oe_1	.472[1]	0	0	.676	0	0	0
Oe_2	.472[1]	0	0	0	.606	0	0
Op	.408	0	0	0	0	.767	0
Omc	.492	0	0	0	0	0	.452
Se_1	0	.535[1]	0	.538	0	0	0
Se_2	0	.535[1]	0	0	.496	0	0
Sp	0	.604	0	0	0	.414	0
Smc	0	.410	0	0	0	0	.475
Me_1	0	0	.768[1]	.290	0	0	0
Me_2	0	0	.768[1]	0	.178	0	0
Mp	0	0	.719	0	0	.197	0

$\lambda^{0}_{83} = 95.547$ ($p \leq 0.164$)

*All estimated coefficients exceed twice their standard errors.
[1]Coefficients fixed at values from Table 10.

FIGURE 1: Path Diagrams for MTMM Factor Analyses of Genre Effects on Essay Data

Panel A



Panel B

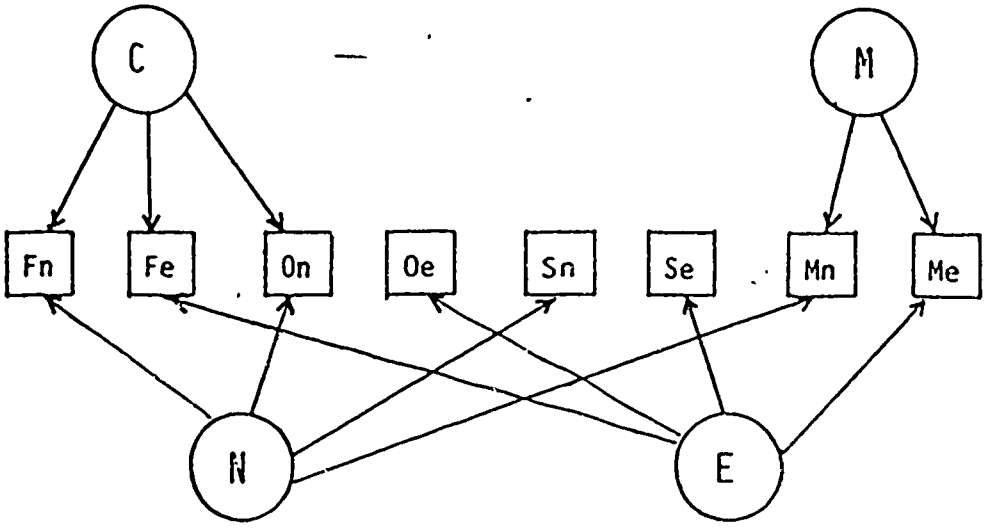
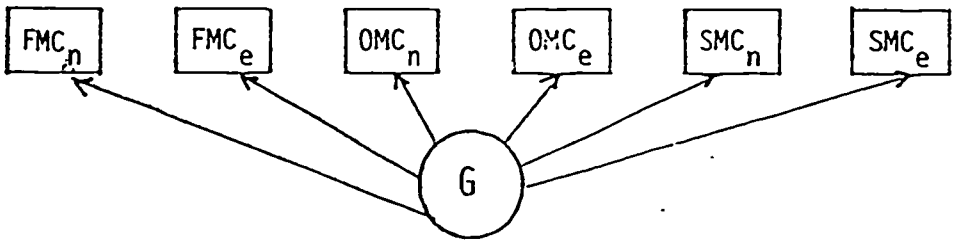
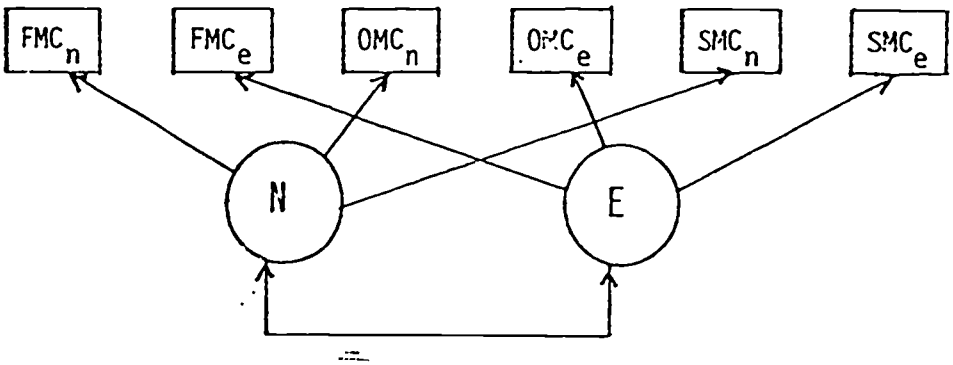


Figure 2: Path Diagrams for MTMM Factor Analyses of Multiple Choice Writing Data

Panel A



Panel B



Panel C

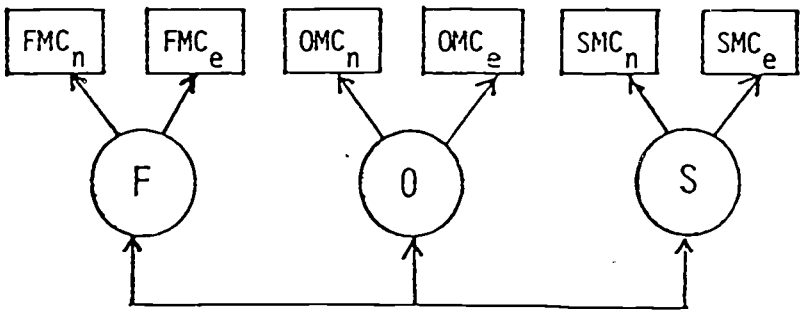


FIGURE 3

Path Diagram for MTMM Model of Response Mode and Trait Effects in Essay Data

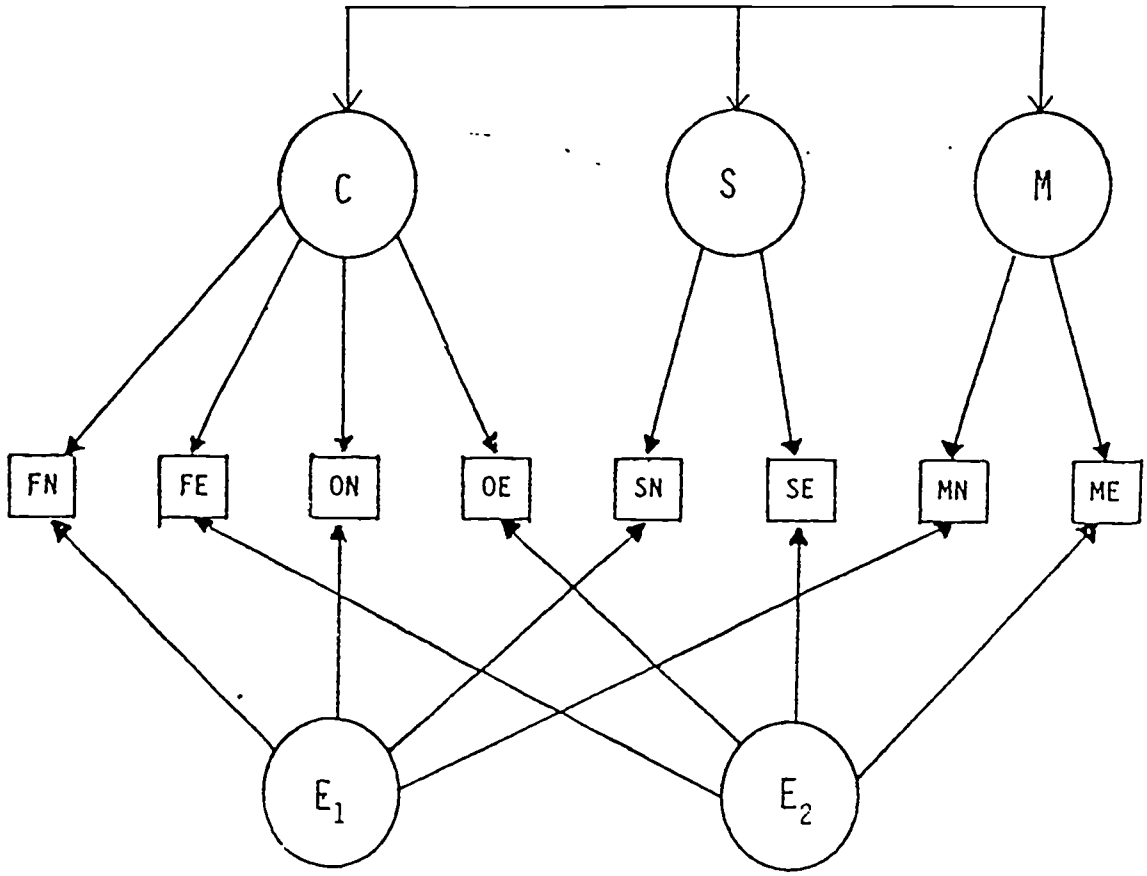


Figure 4

Path Diagram for MTMM Model of Response Mode and Trait Effects in Essay and Paragraph Data

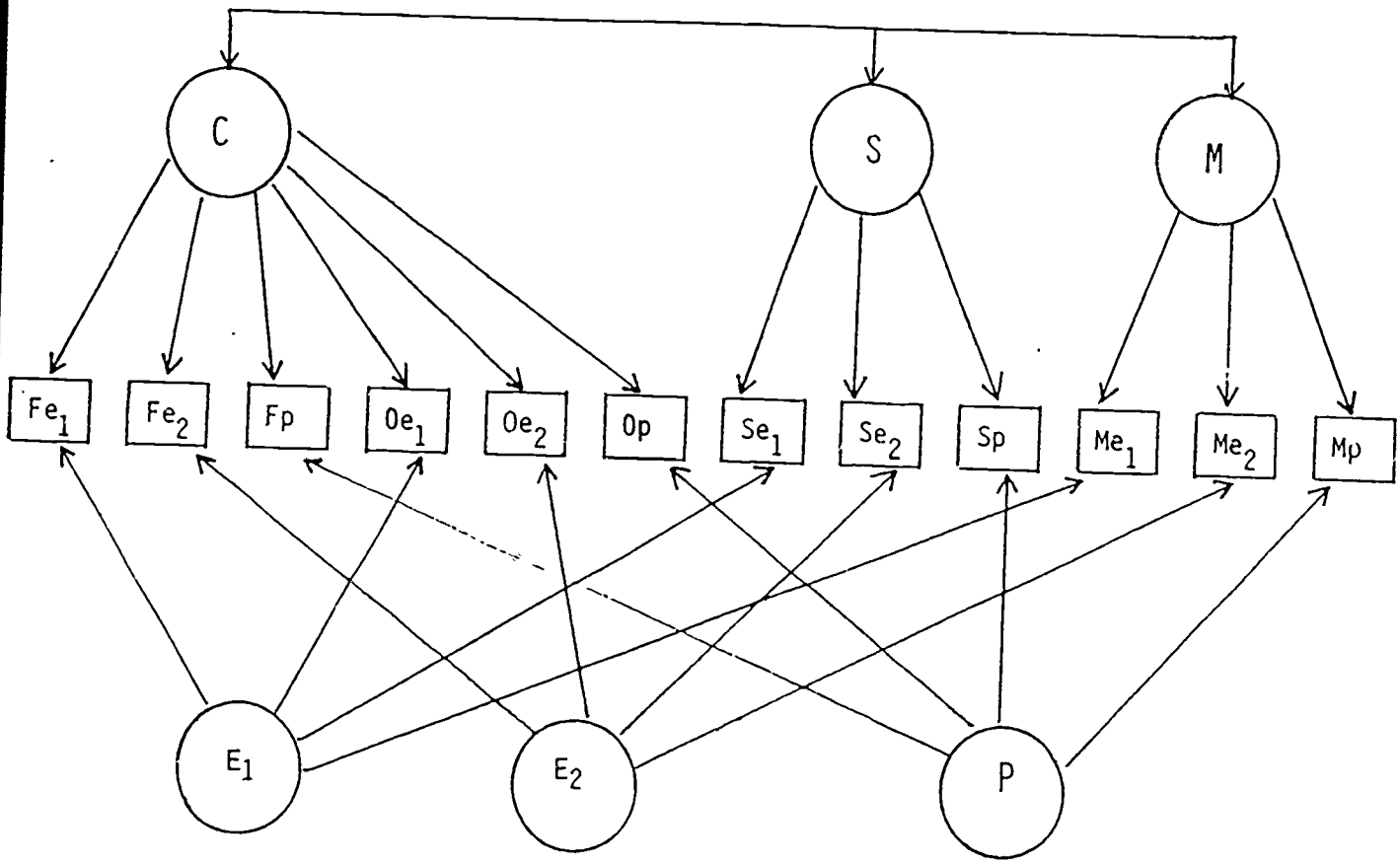


Figure 5

Path Diagram for Models of Response Mode and Trait Effects
in Essay, Paragraph and Multiple Choice Data